

# Mapping the evolution of industrial \*clusters\*: a meta-analysis

Mario A. Maggioni <sup>^</sup>, Francesca Gamarotto <sup>\*</sup>, T. Erika Uberti <sup>^</sup>

<sup>^</sup> DISEIS and Faculty of Political Science, Catholic University, Milano

<sup>\*</sup> University of Padova

## Abstract

This paper presents a meta-analysis of the \*cluster\* literature (here agglomeration of firms and not the statistical tool), contained in scientific journals from 1969 to 2007. Thanks to an original database we study the evolution of a stream of literature which focuses on a research object which is both a theoretical puzzle and an empirical widespread evidence. We study the relationships between a “spatial” and an “industrial” approach within the textual corpus of cluster literature and we measure the extent of a convergence process of the vocabulary of scientist working on \*clusters\*.

**Keywords:** Cluster, life-cycle, cluster literature, textual analysis, agglomeration

## 1. Introduction

In recent years many theoretical and empirical papers have analysed the genesis, development, functioning and decline of \*clusters\*, i.e. agglomeration of firms <sup>1</sup>. This \*cluster\* model has been seized on by public authorities and policy makers as a tool for promoting competitiveness, innovation and growth at local, regional and, sometimes, national level. However, to quote Martin and Sunley (2003: 7), «the mere popularity of a construct is by no means a guarantee of its profundity. Seductive though the cluster concept is, there is much about it that is problematic, and the rush to employ ‘cluster ideas’ has run ahead of many fundamental conceptual, theoretical and empirical questions».

We underline that we are interested in the economic concept of \*cluster\* and that we have disambiguated the word from its statistical use. That is why we tag it with stars in order to not confuse it with the statistical cluster meaning.

This paper looks at \*clusters\* and, in particular, at \*cluster\* life-cycle, from a different perspective. Many papers have been written on the development pattern of industrial \*clusters\*. However since \*clusters\* are a research topic which has been investigated for, at least, 40 years, we think that it is now the right time to look at the evolution (and check whether there is a life-cycle) of \*clusters\* not as economic phenomena, but as research objects. A seminal contribution on this issue is Maskell and Kebir (2006).

---

<sup>1</sup> See, among others, Bresnahan and Gambardella (2004); Breschi and Malerba (2005); Asheim et al. (2006); Braunerhjelm and Feldman (2006); Karlsson (2008).

For this reason, we present a meta-analysis of the \*cluster\* literature (i.e. agglomeration of firms) contained in scientific journals from 1969 to 2007. Thanks to an original database, built for this purpose, we are able to study the evolution of \*clusters\* by studying the development of a stream of literature which is at the cross-roads of different scientific disciplines (economics, geography, environmental studies, regional science, urban studies, business and management, transport studies) and focuses on a research object which is both a theoretical puzzle and an empirical widespread evidence.

The rest of the paper is organised as follows: section 2 describes the selection and construction of the bibliographic database; section 3 studies the evolution of the vocabulary used in this literature; section 4 analyses the existence of \*clustering\* phenomena and dynamics in the \*cluster\* literature with reference to “textual geography” (i.e. based on the specific vocabulary used by scientists in different countries); section 5 analyses inter-country distance in correspondence analysis and the conclusion section follows.

## 2. The database

To analyse the evolution of the concept and use of \*clusters\* in the economic literature we selected articles published in international scientific journals collected in two databases, ISI-Thomson “Web of Science” and “EconLit”, following a three-steps procedure.

Firstly, in December 2007, we downloaded all titles and abstracts (where available) of all articles containing the word \*cluster\*.

In particular, as far as the ISI-Thomson “Web of Science” (henceforth ISI) database is concerned, we limited the analysis to a subset (i.e. economics, planning and development, geography, management, environmental studies, business and urban studies) of all subject categories contained in the social science citation index. In addition we excluded those articles referred to <cluster analysis> as a statistical procedure.

Similarly, we conducted the same textual search in the EconLit database. To avoid any unbalance between Econlit and ISI – and to include information needed to conduct this analysis (i.e. texts of abstracts and affiliations to geo-code authors) – we excluded volumes, collective volume articles, working papers and Ph.D thesis which are contained in the EconLit database, and we limited our search to journal articles.

Secondly, we excluded from our database those articles (such as comments, introductions of special issues, editorials and book reviews) that do not contain original research output and which lack information needed for both textual and geographical location analyses.

Thirdly, we merged both databases in order to avoid duplications so to obtain a final set of 499 scientific articles ranging from 1969, when the first article <sup>2</sup> on \*clusters\* was published, until 2007.

Due to the different purposes of ISI and EconLit databases, we were able to include in our analysis papers published in academic journals with a national and international coverage, hence to capture possible national differences. The degree of overlap between the two databases varies between 13% (in 1996) and 67% (in 1999) while in 1979 and 1989 both databases include the same articles.

---

<sup>2</sup> The first article in the database is “Some properties of a cluster point process”, by M.F. Dacey, and published in the *Canadian Geographer*.

Since we wanted to investigate the existence of \*clustering\* behaviour displayed by people writing on \*clusters\* and, more generally, we wanted to analyse the spatial distribution of this literature, we extracted information from each record included in our final database regarding each author's affiliation<sup>3</sup>. In this way we were able to associate each paper to one (or more) city and country<sup>4</sup>, in order to conduct both a geographical analysis of the \*cluster\* scientific community and a textual analysis focused on national differences and similarities of the vocabulary<sup>5</sup>.

In our database papers were written by authors working in 282 cities located in 45 nations; however, in the "textual geography" analysis (section 4) we focused exclusively on the most relevant countries (Canada, France, Germany, Italy, Netherlands, Spain, Sweden, United Kingdom, United States) which accounted for over 85% of the papers.

Finally, to describe the evolution of the vocabulary used in the titles and abstracts of the articles included in our dataset, we used a "textual analysis" procedure limiting the analysis to the last 10 years (1998-2007), since over 90% of the total scientific production (and of the vocabulary) has been published in these years. As it will be described in greater detail in section 3, the selection of a given initial year is crucial to calculate the IT index.

### 3. The life-cycle of words and concepts

Before building the matrix of graphical forms (*word-types*) we analyzed the quality of the textual corpus using a specific software application, TALTAC2 ([www.taltac.it](http://www.taltac.it)). The corpus dimension,  $N$ , is given by the number of *word tokens* (statistical textual units) while the vocabulary includes  $V$  distinct words, or *word-types*. Each word-type is associated with its number of *word tokens*, i.e. with its frequency. Observing our textual corpus, it results that it is a medium-large corpus including 76.262 word tokens ( $N$ ) while the extracted vocabulary is made up of 6.468 word-types ( $V$ ). Our analysis will be based on the vocabulary ( $V$ ) to identify the main features of the \*cluster\* literature and to interpret its evolution (Lebart et al., 1998; Tuzzi, 2003).

By using the abovementioned software application then we applied the standard procedures to extract textual information from the corpus. The lexicalized vocabulary was defined selecting characteristic words with a *p-value* equals to 0.02 while we did not introduce an exogenous frequency threshold for word-types since we are interested in detecting new conceptual entries in the vocabulary. The lexicalized vocabulary includes 6.807 word-types.

Therefore we calculated characteristic textual units for word-tokens appearing only once. Then, we calculated the IT index (Bolasco and Canzonetti, 2003; Bolasco, 2005), an index used to identify trends and life-cycles of words in the evolution of a given textual corpus over time. To define trends, we selected word-types with a dispersion index between 0,1 and 0,8 ruling out hapax and very frequent words, i.e. the basic vocabulary. We obtained 721 word-types on which it has been possible to apply the following expression:

---

<sup>3</sup> Since our interest was mainly "geographic" in its scope, we limited the analysis to cities and we avoided any investigation on specific universities, centres of research and/or departments.

<sup>4</sup> We extracted city information from each affiliation as defined by authors. In case of multiple affiliations, since in both databases there is no possibility to distinguish between primary and secondary affiliations, we maintained both information.

<sup>5</sup> When authors, based in different countries, write jointly a paper, for the purpose of this textual analysis we attributed the paper to both countries.

$$IT = \left\{ \left[ \prod_{y=2}^n \left[ (Occ_y - Occ_M) * (Occ_{y-1} - Occ_M) \right] \right] / \left[ \prod_{y=2}^n \left[ (Occ_y + Occ_M) * (Occ_{y-1} + Occ_M) \right] \right] - 1 \right\} / 2$$

where, for each word-type, we consider the deviation between the number of normalized occurrences for each year (or word tokens,  $Occ_y$ ) and the equidistributed occurrences (i.e. its mean value,  $Occ_M$ ). The possible values of IT are -1 and 0 and multiplying it by  $(Occ_1 - Occ_n) / (Occ_1 + Occ_n)$  we obtain a IT index varying between -1 and 1. This allows the identification of “extinct” words ( $IT = -1$ ), new words ( $IT = 1$ ), and different typologies of trends <sup>6</sup>. Given that our textual corpus developed significantly during the last ten years (so that we have positive word-tokens at the starting year for around 60% of the 721 word-types) we constrained this analysis (and the calculation of the IT index) only to the period ranging from 1998 to 2007 (see tables 1 and 2 and figure 1 for some examples of different types of trends) <sup>7</sup>.

<i>Trend</i>	<i>Interval</i>	<i>IT value</i>	<i>Word-type</i>	<i>%</i>
Obsolescent		-1	27	3,7
Very decreasing		-1...-0,5	96	13,3
Decreasing		-0,5...0	49	6,8
Bimodal	Occ10 – OccM > 0	0	172	23,9
Unimodal	Occ10 – OccM < 0	0	133	18,4
Growing		0... 0,5	54	7,5
Very growing		0,5 ... 1	9	1,2
Neologism		1	132	18,3
Non identifiable			49	6,8
Total			721	100

*Table 1: Evolutionary trends of word-types (1998-2007)*

Source: our calculations on ISI and EconLit databases

As shown in Tab. 1, most word-types used to build up the theoretical framework of the \*cluster\* concept, show a bimodal or unimodal trends (42%). This means that some words have attracted scholars’ attention in particular years (unimodal), while others present an “up and down” trend suggesting they have irregularly contributed to the scientific debate. We can suppose that these particular trends depend on occasional participation of new scholars to this topic, or derive from the exploration of new theoretical paths.

In addition, Tab. 1 shows that both the very decreasing (13.3%) and neologism (18.3%) trends are very significant for the evolution of the \*cluster\* concept. In our opinion, some word-types are very decreasing because the theoretical focus changed during time, while others have been subsumed in other theoretical categories (e.g. <places> became <economies>). The relevance of the neologism group (18.3%) suggests that the \*cluster\* topic has enriched its lexicon during the period at study.

<sup>6</sup> It should be noticed that trend identification depends on the frequency gap between the limit values of the time period (Bolasco and Canzonetti, 2003).

<sup>7</sup> The stability of the IT index and trends identification crucially depend on the changing limit values. If we chose 1997 as starting point, only the 24% of the selected textual units had a positive frequency at the beginning of period; while opting for 1999, the value increased (83%). A high number of word-types starting with zero frequency and finishing with a positive value are categorized as neologism.

Tab. 2 displays some significant word-types for each group identified in Tab. 1, while graphical examples of the main profiles are portrayed in Fig. 3.

It is worth noting that the evolution of specific word-types must be considered together with other different linguistic “inflections” that the same concept acquired over time. <Evolution> displays a unimodal pattern with a peak in 2006 but <evolutionary> acts as a neologism with an increasing use in the last years; <technological> is bimodal, <technology> decreasing but <technologies> is growing and <biotechnologies> a neologism; <institutions> is bimodal but <context> is growing and <governance> is a neologism.

<i>Characteristic textual elements</i>	<i>IT value</i>	<i>1998</i>	<i>1999</i>	<i>2000</i>	<i>2001</i>	<i>2002</i>	<i>2003</i>	<i>2004</i>	<i>2005</i>	<i>2006</i>	<i>2007</i>
<b>Obsolescent</b>											
developments	-1,00	4	0	0	2	2	8	4	3	0	0
specialization	-1,00	4	5	3	0	11	1	1	3	4	0
distance	-1,00	4	3	3	6	0	0	4	5	2	0
<b>Very decreasing</b>											
quality	-0,78	11	11	3	4	2	2	0	3	1	2
infrastructure	-0,52	11	0	0	2	3	4	3	1	2	1
local firms	-0,52	4	3	0	0	2	6	5	3	2	1
<b>Decreasing</b>											
process	-0,44	15	11	11	16	11	20	15	13	13	6
located	-0,23	4	0	5	4	3	1	0	3	6	2
technology	-0,17	15	11	3	10	18	11	12	13	9	11
<b>Bimodal</b>											
technological	0,00	15	8	11	4	15	7	4	10	12	11
competitive advantage	0,00	11	0	8	2	2	4	8	1	7	7
institutions	0,00	8	0	8	4	5	6	4	9	7	8
<b>Unimodal</b>											
growth	0,00	8	26	38	41	21	16	28	24	28	19
evolution	0,00	0	11	5	4	6	6	9	8	12	4
sector	0,00	8	0	8	16	21	7	14	16	13	5
<b>Growing</b>											
global	0,11	8	11	27	12	24	8	12	15	20	19
results	0,43	11	5	19	10	9	14	8	15	15	18
management	0,48	4	0	3	2	6	1	5	5	2	8
<b>Very growing</b>											
performance	0,52	8	5	14	18	6	6	5	13	18	29
linkages	0,59	4	18	30	6	5	20	16	7	0	12
<b>Neologism</b>											
biotechnology	1,00	0	0	0	51	12	14	12	5	5	13
capabilities	1,00	0	3	8	6	6	2	5	9	4	13
processes	1,00	0	3	11	8	6	7	10	8	4	15

Table 2: A sample of characteristic textual elements (normalised frequencies)

Source: our calculations on ISI and EconLit databases

It is worthwhile to observe that words with the same root evolve differently. For example, we suppose that the different trend of <evolution> and <evolutionary> is due to their different contexts: using the concordance device, we observe that <evolution> underlines the investigation of the \*cluster\* historical path while <evolutionary> refers to a specific economic methodology

that investigates the economic behaviour looking for fitness instead of optimization and reaction capability instead of perfect rationality. This means that this economic paradigm is reinforcing its explanatory role for this phenomenon. As far as the <technolog\*> is concerned, the concordance analysis shows that at the end of the decade, <technology> is more frequently used as a category of <technologies>, especially when the subject is ICT.

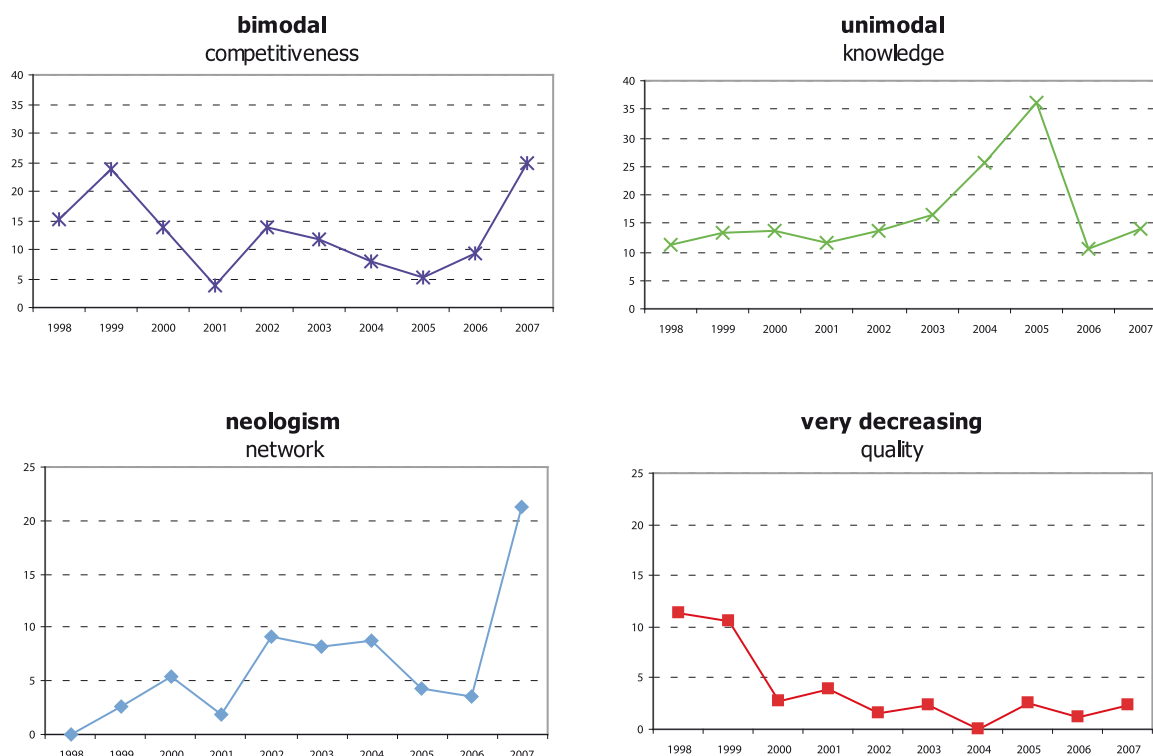


Figure 1: Examples of trends typologies

Source: our calculations on ISI and EconLit databases

#### 4. “Textual” geography

Since we are dealing with a meta-analysis based on a bibliographic database it seemed sensible to further exploit the textual corpus looking for national characteristics “hidden” in the vocabulary used in titles and abstracts.

For this reason we applied correspondence analysis (CA) to the study of the evolution over time of different research communities (scientists belonging to the same countries and participating in the \*cluster\* literature), to detect specialization patterns and the presence of convergence/divergence dynamics. Data matrix was built on the lexicalized vocabulary in the following way: first, we merged plural and singular for crucial words (we reduced to singular/plural following the most common use of words in the literature). Secondly we eliminated both basic vocabulary and words with a frequency less than ten occurrences (first decile). However we saved less frequent but relevant words in the vocabulary. Partitioning the paper’ statistical distribution 1969-2007 in quartile, we obtain enough textual variability for each time period to observe the lexicon changing of the major national scientific communities contributing to the \*cluster\* literature. For AC, we have four matrices (1969-2000; 2001-2003; 2004-2005; 2006-2007) of words (row) for nine countries (column). The software used to run CA is the statistical package SPAD.

Below we present the outcomes of CA for the first two eigenvalues that measure the variances along each principal axis and that, cumulatively, absorb more than the 50% of inertia for each period. Interpretation of the map relies on the positions of the topics with respect the observed countries (research communities). First of all we have interpreted the word clouds (lexicon of research topic) observing the absolute contribution of each word, then we have observed the similarity/difference for countries (column profiles).

In Fig. 2, referring to the first statistical period 1969-2000 (108 papers), the configuration of research topics presents four separated clusters: in the first quadrant scientists investigate the economic *\*cluster\** looking for key factors of local development (e.g. <interdependencies>, <trust>, <networks>, <social>, <collective>) within a “regional science” theoretical framework (<cluster>, <regional>, <local>, <district>). From the cloud in the second quadrant emerges a research path addressing the <life> <cycle> of *\*clusters\** to detect <conditions> for the crucial starting time (<spin-off>) and for *\*cluster\** strengthening (<upgrading>, <practices>). In the third quadrant *\*cluster\** research is managed mixing topics in industrial investigation (<technology> <patents>, <biomedical>), and territorial analysis (<place>, <metropolitan>, <geographic>). The last very scattered cloud is focused on the core terms of this literature: <cluster>, <strategy>, <industrial>, <regional> with a particular attention to elements like <infrastructure>, <resources>, <public>, <disadvantages>. In this first large period, scientist communities ground and sometimes share their interests on specific investigation topics: British scientists (UK), and, by a smaller extent, Italians (ITA) and Spanish (ESP) share the socio-economic perspective focusing on local production systems as engine of economic growth. Their attention is directed to describe the complex network of economic, social and institutional factors that affect the *\*cluster\** stemming from as well as its persistence. The German research community (DEU) develops a very own research topic facing conditions to create the atmosphere, and policies, for a virtuous development cycle for *\*cluster\**. The Sweden (SWE) and France (FRA) communities share the interest for similar topics combining technology and *\*cluster\** economic phenomenon. The American (USA) community has different investigation streams because the technological topic is very stressed as well as questions concerning the *\*cluster\** life-cycle. Canadian (CAN) and Dutch (NLD) scientists community play a less important role in this first research time and are more focused on detecting the infrastructure role for *\*cluster\** development.

In Fig. 3, referring the second statistical period 2001-2003 (134 papers), we have two word clouds that show a research deepening and dominance: the cloud on the right of the figure concerns the analysis of *\*clusters\** in terms of local innovation systems especially investigating the relationships between <university> and <laboratories> and looking for what <competencies> and what territories are more productive. About the cloud on the left, we identify two distinct research approaches even if squeezed: in the second quadrant an institutionalist approach (<subcontractors>, <interdependencies>, <agents>, <community>, <product>) and on the third one a quantitative and formalised one (<autocorrelation>, <assumptions>, <equilibria>). In this period, scientist communities focus the research interests developed previously; the lexicon of the CAN research community is strongly concentrated on the geography of innovation, especially on the production of knowledge spillovers stemming from cooperation between universities and private R&D laboratories. The ITA, NLD, SWE, UK communities, that contribute very little to the definition of the first two eigenvalues, partly share this perspective but they give a particular theoretical and empirical attention to the network features of the industrial *\*clusters\**. The ESP and FRA communities are very rooted on an institutionalist approach and share the interest to look for the social conditions affecting the *\*clustering\** preference of firms. Opposite, from a theoretical





to the vocabulary of this literature. Topics of investigation in this period get deeper to the \*cluster\* role in the knowledge production distinguishing between investigation in behavioural models (first quadrant), evolutionary models (second quadrant) and socio-economic investigations (third and fourth quadrants). Scientific communities foster their scientific contribution deepening their research path; SWE and DEU (third quadrant) share the vocabulary for investigation of both knowledge production (<learning>, <innovation>, <process>, <mobility>) and social communication (<tacit>, <buzz>, <codified>) as well as <labour> market and <mobility>. Moving to the second quadrant, we observe a “cluster” of scientific communities focused on an <evolutionary> perspective: ITA, ESP, CAN, UK but especially the NLD community. For them technology is considered a major driver of cluster development (<ict>, <high-tech>, <technology>) and there is lot of interest in the emergence of co-ordination procedures and structures (<governance>, <upgrading>, <transformation>, <partnership>, <strategic>). The first quadrant hosts only the US community, specialised in quantitative investigation with a strong emphasis on the methodological side (<methodology>, <method>, <methods>, <model>) where the key factors of cluster development are several actors (<university>, <enterprise>, <companies>, <industry>), mobilising capitals (<foreign> <investments>, <financial>, <resources>). FRA community stands alone in the fourth quadrant because it has developed a specific lexicon of investigation concerning the relationships between territory, institutions and social contexts (<coordination>, <relational>, <proximity>, dynamics>, <milieu>).



Figure 4: CA for the third period, 2004-2005  
 Source: our calculations on ISI and EconLit databases

In Fig. 5, referring to the fourth statistical period 2006-2007 (112 papers), we observe that the lexicon to describe and explain the firms’ \*cluster\* is now strengthened and shared. US community share now with DEU and CAN communities the interests for methodological issues (<method>, <statistics>) and for the analysis of the spatial re-allocation of production due to globalisation dynamics (<internationalization>, <multinational>, <mobility>, <strategy>),

<job>, <regional>). NLD and FRA communities share common interests focusing on advanced technologies (<ict>, <high-tech>) and studying those spatial conditions (<economies>, <concentration>, <co-location>) that sustain <co-operation> and produce <efficiency> and <productivity>. ITA, ESP and SWE communities deepen the interaction of industrial and organisational concepts (<district>, <knowledge>, <organization>) in explaining the crucial role played by \*clusters\* and their internal structures of interdependences (<proximity>, <co-operation>, <complementarities>) in shaping the innovation process (<bio-tech>, <diffusion>, <spillover>, <selection>, <smes>). The UK community plays a minor role in the definition of the first two axes of this CA.

Looking at the evolution of these four diachronic CA graphs, it is possible to conclude that, in general, each national scientific community shows a significant degree of continuity in the choice of research topics. The contribution of Canadian scientists is based on one hand on innovation systems and policies, on the other on the process of knowledge diffusion; attention is paid to the role of public authorities in the process of knowledge production, and to the relationships between University and Industry. The research path of US scientists is dominated by the investigation of a quantitative-assiomatic methodology for studying \*clusters\* and \*clustering\*. North-American scholars focus on theoretical models and econometric analyses aimed at interpreting and explaining the performance of \*clusters\* in terms of knowledge production and diffusion. This focus is partially shared by Canadian and German scientists.



Figure 5: CA for the fourth period, 2006-2007

Source: our calculations on ISI and EconLit databases

The existence of a common “European Research Area”<sup>8</sup> does not appear to have shaped the research activity of scholars contributing to the \*cluster\* literature. ITA and ESP communities

<sup>8</sup> To quote the title of the famous communication of the European commission (European Commission, 2000).

deepened their investigation over time on the basis of the Marshallian concept of industrial district.

The FRA community has developed a specific research model, rooted in Perroux's thought. Analyses are focused on the relationships between space and technology (e.g. growth poles) assuming, as basic ingredients, cognitive and institutional dimensions. In the first period the NLD community showed some similarities with the CAN one but, more recently, it moved its attention to knowledge production and regional evolution. The SWE community appears to be very focused on knowledge production and diffusion while the UK one combines the interest for both neo-classical and institutionalist approaches. The DEU community is the European country closer to the USA one with respect to research interests; however, it also cultivates research questions, more similar to the SWE scientist community, concerning the spatial dynamics of local economies and the dynamics of knowledge production and diffusion.

In conclusion, CA shows that specialization and differentiation dynamics in the choice of research trajectories by different national scientist communities coexist. Research communities build their own social preferences (qualitative vs. quantitative methods; neoclassical vs. institutionalist-evolutionary schools; cognitive vs. behavioural approaches, etc.) and their dynamic is strictly connected with these theoretical choices.

However, at the same time, scientists are often called to confront different (if not opposite) positions in workshops, conferences and, more rarely, journals and, from these interactions, the robustness of the analyses is increased and theoretical concepts are further refined. This method of comparison of different ideas and knowledge diffusion process produces, in some cases, a strong homogenization of research topics (which is labelled, in our empirical exercise, as "convergence of research interest") as well as a segregation of original thoughts which cannot be easily integrated in the dominant paradigm. This process may be further re-enforced by the oligopolistic structure of the demand side of the "research market" (i.e the academic journals), which is increasing over time.

## 5. Conclusion

This paper, based on an original database, developed by the authors on the basis of two major independent bibliographic databases (ISI and EconLit), contains a meta-analysis of the \*cluster\* literature in order to describe the evolution of the concept of \*cluster\* in major academic journals and to identify the existence of trends and life-cycles in the textual corpus.

In particular, we identified different phases of growth of this literature, from the birth of the concept to the take off phase, from the development to the current maturity. The IT index allowed us to distinguish between different trends in the use of specific word-types.

We performed a "textual geographic" exercise searching for country-specific vocabulary in the \*cluster\* literature and testing the hypothesis of a convergence dynamic acting over time.

This paper is also a research agenda for the authors. Several interesting issues have been briefly analysed while they would deserve specific in depth analyses. We are convinced that through meta-analyses one could gather in depth knowledge of a given stream of literature, i.e. its nature, history and evolution.

## References

- Asheim B., Cooke P. and Martin R. (editors) (2006). *Clusters and Regional Development. Critical Reflections and Explorations*. London: Routledge.
- Bathelt H., Malmberg A. and Maskell P. (2004). Clusters and Knowledge: Local Buzz, Global Pipelines and the Process of Knowledge Creation. *Progress in Human Geography*, 28: 31-56.
- Bolasco S. (1999). *Analisi multidimensionale dei dati. Metodi, strategie e criteri d'interpretazione*. Roma: Carocci.
- Bolasco S. (2005). Statistica testuale e text mining: alcuni paradigmi applicativi. *Quaderni di Statistica*, vol. 7: 17-53.
- Bolasco S. and Canzonetti A. (2003). Il text mining come chiave di lettura del linguaggio degli anni Novanta su "La Repubblica". In *Convegno Scientifico S.I.S. 2003 sul tema: Analisi statistica multivariate per le Scienze Economico-Sociali, le Scienze Naturali e la Tecnologia*, Università di Napoli "Federico II", 9-11 giugno.
- Braunerhjelm P. and Feldman M. (editors) (2006). *Cluster Genesis. Technology-based Industrial Development*. Oxford: Oxford University Press.
- Breschi S. and Melerba F. (editors) (2005). *Clusters, Networks, and Innovation*. Oxford: Oxford University Press.
- Bresnahan T.F. and Gambardella A. (editors) (2004). *Building high-tech clusters: Silicon Valley and beyond*. Cambridge: Cambridge University Press.
- Duriau V., Reger R.K. and Pfarrer M.D. (2007). A Content Analysis of the Content Analysis Literature in Organization Studies- Research Themes, Data Sources, and Methodological Refinements. *Organizational Research Methods*, 10: 5-34.
- European Commission (2000). <http://europa.eu/scadplus/leg/en/lvb/i23010.htm>, accessed march 2009.
- Karlsson C. (editor) (2008). *Handbook of Research on Clusters: Theories, Policies and Case studies*. Cheltenham: Edward Elgar.
- Lebart L., Salem A. and Berry L. (1998). *Exploring Textual Data*. Dordrecht: Kluwer Academic Press.
- Martin R. and Sunley P. (2003). Deconstructing clusters: chaotic concept or policy panacea?. *Journal of Economic Geography*, 3: 5-35.
- Maskell P. and Kebir L. (2006). What Qualifies as a Cluster Theory?. In Asheim, B., Cooke, P. and Martin, R., editors, *Clusters and Regional Development. Critical Reflections and Explorations*, London: Routledge, pp. 30-49.
- Tuzzi A. (2003). *L'Analisi del Contenuto. Introduzione ai metodi e alle tecniche di ricerca*. Roma: Carocci.